Optimal Scheduling of the Leaves of a Tree and the SVO Frequencies of Languages

Christos H. Papadimitriou and Denis Turcu(✉)

Columbia University, New York, NY 10027, USA
{christos,d.turcu}@columbia.edu

Abstract. We define and study algorithmically a novel optimization problem related to the sequential scheduling of the leaves of a binary tree in a given order, and its generalization in which the optimum order is sought. We assume that the scheduling process starts at the root of the tree and continues breadth-first in parallel, albeit with possible intervening lock and unlock steps, which define the scheduling cost. The motivation for this problem comes from modeling language generation in the brain. We show that optimality considerations in this problem provide a new explanation for an intriguing phenomenon in linguistics, namely that certain ways of ordering the subject, verb, and object in a sentence are far more common in world languages than others.

Keywords: Binary tree · Optimal leaves scheduling · Language · Basic word order

1 Introduction: The Leaf Scheduling Problem

Consider a binary tree T, e.g. the one in Fig. 1A—where by binary tree we mean a downwards directed tree with $2n-1$ nodes, one node of degree two (the root), $n-2$ nodes of degree three, and n nodes (the leaves) of degree one—and suppose that we are also given an order σ of the leaves, say the order subject-verb-object (SVO) in this example. We are interested in assigning integer times to the nodes of the tree according to the following rules:

1. The root is assigned time 0;
2. A non-root node i either is assigned time $t+1$, where t is the time assigned to its parent, or it is *locked* by its parent;
3. The leaves are assigned times that are strictly increasing in the given order, σ;
4. If a leaf ℓ is assigned a time t, then a locked node i may be assigned time $t+1$, in which case we say that ℓ *unlocks* i.

We say that an assignment is a *feasible schedule* if it satisfies these rules. Intuitively, this assignment of times formalizes the process in which the nodes of the tree "fire" starting with the root, and the children of a node fire right

D. E. Simos et al. (Eds.): LION 2022, LNCS 13621, pp. 3–14, 2022.
https://doi.org/10.1007/978-3-031-24866-5_1

after their parent did. The exception is that a node may choose to lock one or both of its children at the time of its own firing. A leaf, upon firing, may unlock one locked node. It is clear that, given a tree and an order of its leaves, there is always a feasible schedule: Always lock the child that does not lead to the next leaf in the order, while any firing leaf unlocks the locked ancestor of the next leaf in the order.

Define now the Leaf Scheduling Problem to be following: *Given a tree T and an order σ of its leaves, find a feasible schedule that has the smallest cost, where the cost of a schedule is the number of lock commands used (equivalently, the smallest number of unlock commands).* We can also define the *weighted version* of this problem by assigning a weight to every possible lock and unlock command, and minimizing the sum of these weights. We can further define a more complex problem called the *optimum leaf order problem*, in which we are only given a tree T and we seek the order σ that has the smallest scheduling cost.

For example, the optimum leaf scheduling problem for the order SVO in Fig. 1A is the one that assigns 0 to the root, 1 to S and the internal node, 2 to V and 3 to O. That is, the internal node locks O, and V unlocks it. This solution has cost one, since one lock is used, and it is clear that there is no solution with zero cost. In fact, the order S–V–O along with S–O–V are the optimum leaf orders with cost one, while the other four orders have optimum cost two. For a more complicated example, the reader may want to verify that the tree in Fig. 1B, with the leaf order from left to right, has optimum leaf scheduling cost three, while the optimum orders for this tree are the orders ADBC and ADCB with cost one. As we shall see in the next section, both algorithmic problems can be solved by greedy algorithms — with the exception of the optimum leaf order problem with weights, which is NP-hard.

Motivation: Word Orders in Natural Languages

The reason these problems are interesting is because they relate to a classical problem in Linguistics, which we explain next. In English, the subject of a sentence generally comes before the verb while the object, if present, follows both: "dogs chase cats". This ordering is not universal, as other languages adopt any of the six possible orderings, see for example [4]. The same order as in English, denoted SVO, is prevalent in French, Hebrew, modern Greek and Romanian, and overall in about 42% of world languages. The order SOV is slightly more common, accounting for 45% of languages, including Hindi, Urdu, Japanese, Latin, and ancient Greek. The orders VSO (9%), VOS (2%) and OVS (1%) are much less common, while the order OSV ($< 1\%$) is practically disregarded. In English, changing the language's SVO order creates either meaningless sentences ("chase cats dogs") or changes the meaning ("cats chase dogs"). In other languages, such as German, Russian, or modern Greek, deviations from the standard order are tolerated, because nouns have a *case* in these languages, which makes their syntactic role (subject *vs* object) easy to identify independently of position. However, many linguists believe that most languages have a dominant, default word order.

There is extensive literature on justifying the widely varying frequencies of basic word orders, see [1,8,10,11,14,17–19,22]. These past explanations are based on plausible linguistic principles related to the ease of communicating meaning, or the difficulty of learning grammar [8,9,14,22] while more recent explanations consider the mutability and evolution of word orders in languages [17–19]. Here we propose a different explanation based on *the difficulty of articulating sentences in the brain.*

Indeed, one can hypothesize that, in order to generate a sentence such as "cats chase dogs," a speaker must first create, through neuronal circuits in their brain, a tree representation of the sentence as in Fig. 1A,C. There is cognitive evidence [6] suggesting that this tree is binary (that is, there are no nodes with more than two children), and in fact that the three leaves"cats," "chase" and "dogs" are organized as shown in Fig. 1A,C (instead of the alternative, e.g., where S and V are combined first); see Fig. 3 in [20]. Given now this tree, the speaker must articulate it, and this involves selecting and implementing one of the six word orders. To arrive at one of the orders, a neural mechanism of "lock" and

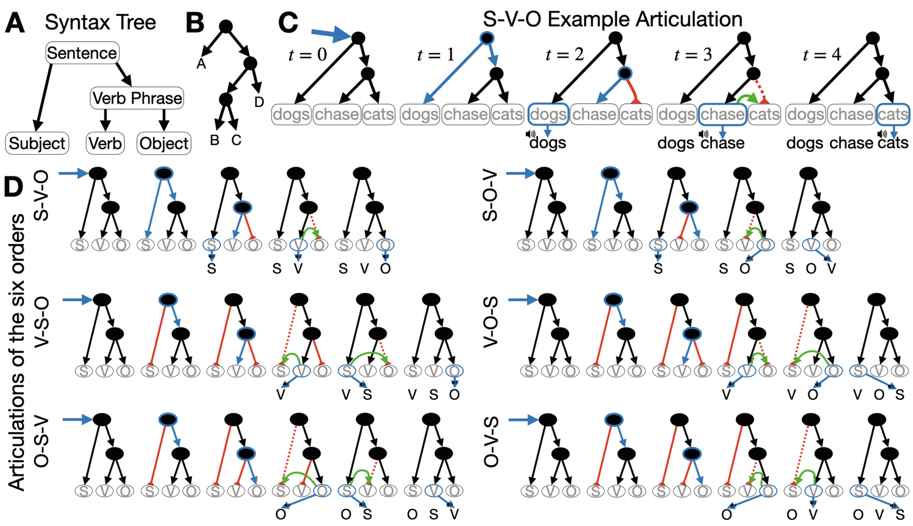

Fig. 1. **(A)** Basic syntactic tree with a "Verb Phrase" internal node and "Subject", "Verb" and "Object" leaves. **(B)** An alternative binary tree example, more complex than the basic syntactic tree. **(C)** Example articulation from the syntactic tree to sequential speech for the SVO order. Black arrows are inactive. Blue arrows activate the object they point to on the next time step. Red inhibitory signals maintain a lock on the object they point to. Green arrows remove the lock. **(D)** Articulations of all possible basic word orders to sequential speech, starting from the same syntactic tree. Appropriate lock and unlock operations dictate the basic word order. (Color figure online)

"unlock" steps may be used. In Sect. 3 we point out that this can easily be done in the model of brain computation proposed in [20]. It would make sense that all speakers of a language end up using the same fixed order, for reasons of effective information transfer; even though most languages allow, unlike English, more flexibility in articulation orders, many linguists believe that there is a dominant order in each language [5,8,17,19,22]. But which of the six orders will be chosen as the dominant order? We propose that, the smaller the implementation cost of a word order in the brain, the more likely it should be for the order to be chosen. In the model of brain computation articulated in [20] and explained further in Sect. 3, every node of the tree resides in a different brain area, and long-range inhibitory neurons are used to lock and unlock brain areas and thus articulate the sentence.

The rest of this paper is organized as follows: In Sect. 2 we study algorithmically the leaf ordering problems, while in Sect. 3 we spell out the model of brain computation in [20] and the way it can implement sentence generation. This model ends up providing an explanation for the differences in the probability of word orders: The two most frequent word orders correspond to the two optimal solutions, while the other four lag behind. Adopting a model in the style of statistical mechanics for calculating the frequencies of the orders allows one to even predict the various differences in the cost of the various lock and unlock steps that would best explain these frequencies.

2 The Greedy Algorithm

Recall the two problems defined in the introduction: The Leaf Sequencing Problem seeks the optimum scheduling of lock and unlock steps that realizes a given sequence, whereas the Optimum Order Problem wants to find the order that minimizes this optimum cost. Both problems can be weighted.

Theorem 1.

1. The Leaf Sequencing Problem can be solved in $O(n \log n)$ time through a greedy algorithm; ditto for the weighted case.
2. The Optimum Order Problem can be solved by an adaptation of the same greedy algorithm, if all lock and unlock steps have unit cost.
3. However, if the unlock steps have different costs, even if the costs are restricted to be either one or two, the Optimum Order Problem is NP-hard.

Proof. (1) We describe the algorithm informally. It entails the sequential firing of all nodes of the tree, starting from the root; the firing propagates from a node to its children down the tree (a breadth-first search implemented by a queue of nodes). Specifically, the root fires at the first parallel time step. At step $t+1$, the internal nodes whose parents fired at step t will fire. Additionally, *any leaf unlocked* by another leaf at time t will fire at time $t+1$ (there will be at most one unlocked leaf at any time step). Finally, if one of the internal nodes firing has any leaf children, then each child is locked *unless it is the next leaf to be output.*

To keep track of leaves we maintain a separate heap of *locked leaves* ordered by σ, initially empty, and an index `next`, initially `1`. If at some step we encounter a leaf child i of a node being processed, there are two cases: If $\sigma(i) = \texttt{next}$, and no other leaf has been output during this step, then the leaf is output immediately and `next` is increased by `1`. Otherwise, $\sigma(i) > \texttt{next}$, and i joins the heap of locked leaves. At the beginning of parallel step t (the round of breadth-first-search processing the nodes of the tree at depth $t-1$), we check whether the `min` of the heap, call it m, has $\sigma(m) = \texttt{next}$. If so, then we output m and increment `next`. We then proceed with the breadth-first search. The algorithm terminates when both the heap and the queue are empty.

We claim that this algorithm outputs the leaves in the σ order, and that it does so with the fewest lock and unlock operations and in the fewest parallel steps possible. We first claim that every leaf is output as early, in terms of parallel time, as possible. This follows from two things: (a) no leaf i can be output earlier than time $T(i)$, where $T(i)$ satisfies the recurrence $T(i) = \max\{T(\sigma^{-1}(\sigma(i) - 1) + 1, \text{depth}(i)\}$ if i is not the first leaf and $T(i) = \text{depth}(i)$ otherwise; and (b) the algorithm achieves this time, as can be shown by induction on $\sigma(i)$. We also claim that it implements the permutation with the fewest locks, which follows from the two facts that (c) the minimum possible number of locks is $n-1$ minus the number of *coincidences,* where a coincidence is an i for which the two terms in the recursive definition of T above are equal, and (d) such coincidences are caught and exploited by the algorithm.

(2) For the Optimum Order Problem, we start by noticing that every leaf i becomes available to be output at time depth(i). Second, a leaf can be output without lock/unlock steps only if it is output at the precise time it becomes available. Otherwise, if many leaves have the same depth, all but one of them can be feasibly postponed to any time in the future, and unlocked by the leaf that was output immediately before it. Hence the following greedy algorithm achieves the minimum number of lock/unlock steps: We define a one-to one mapping from the n leaves to the time slots $\{d, d+1, \ldots, D+n\}$, where d and D is the minimum and maximum depth of a leaf of the tree: First, each leaf i is mapped to depth(i), which creates a map which is not one-to-one because of collisions. We then repeatedly go through the time slots, from smaller to larger starting from d and execute the following algorithm: for any time slot t, if it has $\ell > 1$ leaves mapped to it, select $\ell - 1$ of these leaves and assign them to the $\ell - 1$ empty time slots greater than t and closest to t, resolving ties arbitrarily. It is easy to see that this algorithm chooses the permutation of the leaves which has the maximum number of coincidences (leaves fire exactly when they become available), in the sense of the previous paragraph, and thus the minimum possible number of lock and unlock steps.

(3) Finally, for NP-hardness: Imagine that the tree is a full binary tree of depth d — that is, $n = 2^d$ and all leaves arrive simultaneously. Then all permutations are available, and we need to chose the ones that order $\sigma(1), \sigma(2), \ldots, \sigma(n)$ such that $\sum_{i=2}^{n} \text{unlockcost}(\sigma(i-1), \sigma(i))$ is as small as possible. It is easy to see that this is a generic instance of the (open-loop) traveling salesman problem,

which is known to be NP-hard even if the lengths of the edges are either one or two [21]. This completes the proof of Part (3) and of the theorem.

3 Generating Sentences in the Brain

It is by now widely accepted among neuroscientists that, in the brain, information items such as objects, ideas, words, episodes, etc. are represented by large populations of spiking neurons. These populations are called *assemblies*. In [20], a computational system was presented whose basic data item is the assembly of neurons, and its operations include *merge,* the creation of an assembly that has strong synaptic connectivity to and from two already existing assemblies, as well as operations that inhibit and disinhibit brain areas. Notice that by repeated application of the merge operation, trees can be built. Indeed, a simple sentence such as "dogs chase cats" can be generated by first identifying the three assemblies corresponding to the three words in the lexicon — believed to reside in the left medial temporal lobe [7]. Then, these word-assemblies project to create three new assemblies within separate subareas of Wernicke's area in the superior temporal gyrus, corresponding to Subject, Verb and Object brain areas. Next, the Verb and Object assemblies (in this example corresponding to "chase" and "cats", respectively) merge to create a *Verb Phrase* assembly in Broca's area [7]. Finally, the Subject and Verb Phrase assemblies merge to create an assembly representing the Sentence Fig. 1A, in another subarea of Broca's area [7]. A sentence may have many other constituents, such as determiners, adjectives, adverbs, and propositional phrases, but here we focus only on the tree built from its three basic syntactic parts: Subject, Verb, and Object.

Three different binary trees can be built from three leaves, by grouping any two of these leaves first. There is a broad consensus in Linguistics [5,14,17,22], as well as evidence from cognitive experiments [6], supporting the basic tree described above Fig. 1A with an internal Verb Phrase node whose constituents are Verb and Object.

Once the sentence is generated, it may be *articulated*, that is, converted into speech. This can be done by exciting the root of the tree – the Sentence assembly – which then will excite its children in the tree and so on. Eventually, all three leaves will be excited. Each leaf can mobilize motor programs which will articulate each word, but this must be done sequentially. Therefore, *one of the six orders must be selected and implemented.* Perhaps the simplest and most biologically realistic mechanism for implementing a particular order involves two plausible primitives, which we call *lock* and *unlock*. These primitives correspond to the familiar neural processes of inhibition of an area (the activation of a population of inhibitory neurons which will prevent excitatory neurons in this area from firing) and dis-inhibition (the inhibition of the inhibitory population) [3,12,16]. In particular, upon firing, an assembly in the tree can inhibit one of its children from firing. Secondly, any leaf can, upon firing, dis-inhibit any other leaf.

3.1 Scheduling Cost Explains SVO Frequencies

We have already seen that, among the six orders, only two can be implemented by just one lock and one unlock operation, whereas all others require two lock and two unlock operations Fig. 1D. In other words, this simple model immediately predicts "the highest-order bit" of the frequency statistics, namely the prevalence of the SVO and SOV orders. All other orders besides these two require extra inhibition and disinhibition, primitives that are known to require significant brain energy consumption [2,3,15]. Furthermore, extra operations makes the articulation process more complex, and presumably renders this aspect of language more difficult for the learner.

3.2 Leaf Scheduling Cost as Energy

It has been argued in the literature [5,8,17–19] that languages have undergone transitions in their history, in which the word order has changed, and hence the current frequencies reflect a dynamic equilibrium of this dynamic process. This view motivates a naïve statistical-mechanical formulation, treating the frequencies of the basic word orders as a Boltzmann distribution [13], in which states with energy level L are prevalent with probability proportional to $e^{-\beta L}$, for a temperature parameter β. For simplicity, we take $\beta = 1$ in this account (but in our experiments we use a wide range of values for β Fig. 2). The states of our model are the six basic word orders and the associated energies are the number of operations required by each articulation choice. The optimal choices for SVO or SOV have low energies, requiring only two operations (one lock and one unlock), while the other four optimal choices have high energies, requiring four operations (two lock and two unlock) Fig. 1D. The prevalence of the six orders SVO, SOV, VSO, VOS, OSV, and OVS would be proportional to the numbers $e^{-2}, e^{-2}, e^{-4}, e^{-4}, e^{-4}, e^{-4}$, respectively. The orders SVO and SOV would then be expected to be more frequent than the rest by a factor of $e^2 \approx 7.4$, predicting frequencies $(.39, .39, .055, .055, .055, .055)$, a great first-order approximation of the empirical distribution $(.45, .42, .09..02, .01, .01)$.

The true cost of a brain area locking and unlocking another area may differ, depending on the distance between the two brain areas involved and the strength of their neural connections, as well as the duration, in steps, of the locking state of the target area. By introducing such hyper-parameters, in addition to β, and fitting them to the observed data, we can in fact *predict* their values. That is, make predictions about the connectivity, via inhibitory neural connections, between brain areas. It turns out that these predictions are robust to various hyper-parameters, including β. The Boltzmann distribution model provides the basis for estimating the frequency of the six basic word orders. We equate these frequencies [Equation 2] with the empirical observations and numerically solve the system of six non-linear equations. We note that the equations display an analytical degeneracy which is also recovered from the simulations; specifically, four of the six parameters can only be determined up to a common additive constant. This degeneracy is manifest in Eq. 1, in that these four parameters cannot be compared with the other two.

The system of equations does not have an analytical solution, but the six parameters can be approximated using gradient optimization. This method finds the same qualitative results for different values of the coefficient β Fig. 2A and for different values of other hyper-parameters. The results of these calculations are robust enough to support certain predictions about the relative costs of dis-inhibiting one brain area from another Fig. 2B. More specifically, we find that:

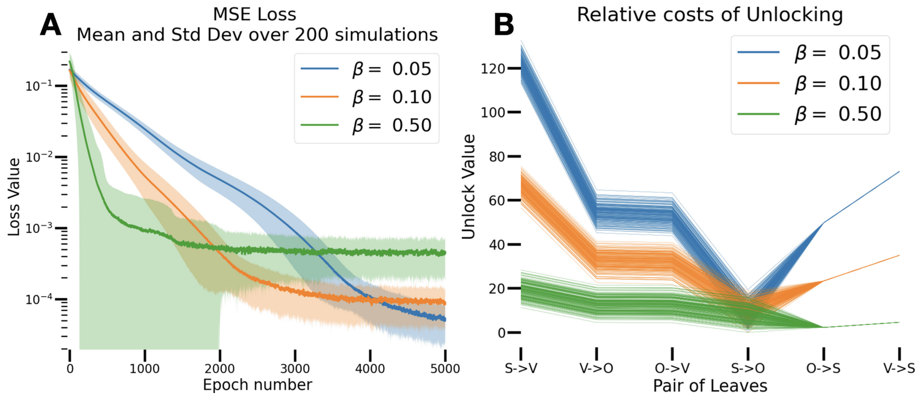

Fig. 2. **(A)** Loss value plotted during the epochs of gradient descent for various β values. The lines represent the average loss and the shaded areas the standard deviation over 200 initializations. **(B)** Relative costs of the unlock operation from one leaf to another. Colors represent models with varying β. Each line represents the optimized parameters for one model. Note the degeneracy of the solutions for the first four leaf pairs: the lines differ only by an additive constant.

$$U_{S\to V} > U_{V\to O} \gtrapprox U_{O\to V} > U_{S\to O} \quad \text{and} \quad U_{V\to S} > U_{O\to S}, \tag{1}$$

where $U_{x\to y}$ is the cost to dis-inhibit assembly y from assembly x.

3.3 A Statistical-Mechanical Argument

In statistical mechanics, the probability of a given state of a system depends on its energy and temperature parameter. The Boltzmann distribution provides a way to estimate the thermal equilibrium configuration of all the states of a system. The probability of a state with energy E_i is proportional to $p_i \propto \exp(-\beta E_i)$, where β is a scale factor, inversely proportional to temperature, and E_i depends on the respective unlock costs $U_{x\to y}$ of state i. The system we describe has six states, therefore, the probability of each state is:

$$p_i = \frac{\exp(-\beta E_i)}{\sum_j \exp(-\beta E_j)}, \tag{2}$$

for $i, j \in \{SVO, SOV, VSO, VOS, OVS, OSV\}$.

These formulas are simply a heuristic way, aligned with physical principles, of modeling how complexity affects probabilities; however, we also note that in the neuroscience literature (see e.g. [15]) metabolic costs are thoroughly discussed with respect to thermal energy. On this account, we choose the energies of our states to be $E_i \sim 1$ and we assume $\beta \lesssim 1$.

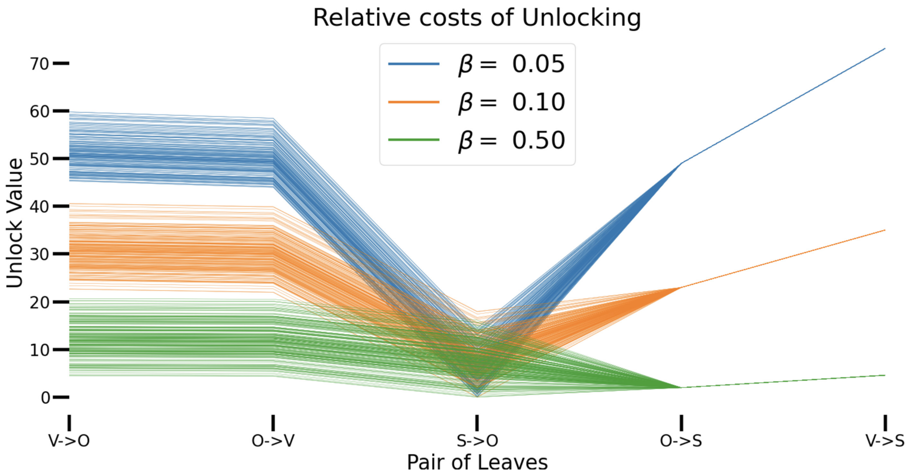

Fig. 3. The predictions if the primacy of Subject and Verb is the cause of the low frequencies of the OSV and OVS orders. Notice that there is no prediction for $U_{S \to V}$.

4 Discussion

Linguistic phenomena should be constantly reinterpreted under the light of new insights, including advancements in our understanding of, or theories about, language processing in the brain. Despite recent progress in this front, articulating the constraints imposed by the neural processes involved in the language function is not easy, due to a large gap, in both scale and focus, between cognitive and systems neuroscience. Our work attempts to bridge this gap using the computational framework of the Assembly Calculus, thus providing a new explanation of the difference in frequencies of the six basic word orders in languages in terms of the difficulty of generating an order from the basic syntax tree of the sentence.

The simplest version of our model qualitatively matches the observed basic word order frequencies, and the most complex version can be tuned to predict the exact frequencies. However, we suspect that the latter calculation may constitute overfitting, as other considerations are likely to enter in the determination of these frequencies, including linguistic considerations of communication efficiency and learnability. These other factors were heretofore the only ones used for this purpose. Our model is not meant to replace these arguments, but add to them and it provides an additional basis for breaking the symmetry of the basic word orders.

We believe that the ultimate explanation of the phenomenon of word orders will integrate both linguistic and neurocomputational evidence, and perhaps learnability considerations, together with more kinds to come. For an example of how this can be done, let us take the linguistic argument that the primary cause of the extreme rarity of orders starting with "O" may not be the difficulty of unlocking Subject or Verb subareas from the Object area according to our model, but the relatively subsidiary semantic role of Object in a sentence, compared to the primacy of the Subject and the Verb [14,17,22]. In the face of this, we may decide that the low frequencies of the OSV and OVS orders are adequately explained on linguistic grounds, and focus on explaining the remaining four frequencies through the corresponding equations. This leads to 4 equations with 5 parameters (since $U_{S\to V}$ no longer enters the picture). To balance the number of equations and parameters, we may fix the ratio of the parameters $U_{O\to S}$ and $U_{V\to S}$ (the ones that were not subject to degeneracy in Fig. 2), and solve by gradient descent. The results are shown in Fig. 3. We notice that our predictions that $U_{V\to O} \gtrapprox U_{O\to V} > U_{S\to O}$ and $U_{V\to S} > U_{O\to S}$ are stable, while our previous prediction that $U_{S\to V}$ is very large vanishes because this unlock operation only plays a role in the OSV order, whose frequency we are ignoring. In other words, the prediction that "$U_{S\to V}$ is very large" was proposed as the cause of the small frequencies of OSV and OVS, a phenomenon which now has another causal explanation based on linguistic principles. It may still be that $U_{S\to V}$ would be large, presumably because this brain connection is rarely used, but a different model or experimental evidence may need to be employed to settle this.

References

1. Aitchison, J.: Word order universals. John A. Hawkins Quantitative Analysis of Linguistic Structure Series. Academic Press, New York, London, Sydney, 1983–1986. https://doi.org/10.1016/0024-3841(86)90039-2
2. Attwell, D., Laughlin, S.B.: An energy budget for signaling in the grey matter of the brain. J. Cerebral Blood Flow Metab. **21**(10), 1133–1145 (2001). https://doi.org/10.1097/00004647-200110000-00001. pMID: 11598490
3. Buzsáki, G., Kaila, K., Raichle, M.: Inhibition and brain work Neuron **56**(5), 771–83 (2007). https://doi.org/10.1016/j.neuron.2007.11.008, https://www.sciencedirect.com/science/article/pii/S0896627307009270
4. Dryer, M.: Order of subject, object, and verb. In: Martin, H., Matthew S.D., David, G., Bernard, C. (eds.) The World Atlas Language Structures, chap. 81, pp. 330–333. Oxford University Press, Oxford (2005)
5. Dunn, M., Greenhill, S.J., Levinson, S.C., Gray, R.D.: Evolved structure of language shows lineage-specific trends in word-order universals. Nature **473**(7345), 79–82 (2011). https://doi.org/10.1038/nature09923

6. Frankland, S.M., Greene, J.D.: Two ways to build a thought: distinct forms of compositional semantic representation across brain regions. Cereb. Cortex **30**(6), 3838–3855 (2020). https://doi.org/10.1093/cercor/bhaa001
7. Friederici, A., Chomsky, N.: Language in Our Brain: the Origins of a Uniquely Human Capacity. The MIT Press (2017). https://books.google.com/books?id=MJg-DwAAQBAJ
8. Gell-Mann, M., Ruhlen, M.: The origin and evolution of word order. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 42, pp. 17290–17295 (2011). https://doi.org/10.1073/pnas.1113716108, http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl.Thisarticlecontainssupportinginformationonlineatwww.pnas.org/lookup/suppl/, https://doi.org/10.1073/pnas.1113716108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1113716108
9. Greenberg, J.: Some universals of grammar with particular reference to the order of meaningful elements. Tech. Rep. 40–70 (1963). http://lear.unive.it//handle/10278/3011
10. Hammarström, H.: The Basic Word Order Typology: an Exhaustive Study. Tech. Rep. (2015)
11. Huber, M., Consortium, T.A.: Order of subject, object, and verb. Atlas of Pidgin and Creole Language Structures Online, pp. 330–333 (2013). http://apics-online.info/parameters/1
12. Jackson, G.M., Draper, A., Dyke, K., Pépés, S.E., Jackson, S.R.: Inhibition, disinhibition, and the control of action in tourette syndrome. Trends in Cognitive Sciences **19**(11), 655–665 (2015). https://doi.org/10.1016/j.tics.2015.08.006, https://www.sciencedirect.com/science/article/pii/S1364661315001837
13. James, S.: Statistical Mechanics: Entropy, Order Parameters and Complexity. In: Oxford Master Series in Physics, OUP Oxford 14 (2006). https://ezproxy.cul.columbia.edu/login?qurl=https
14. Krupa, V.: Syntactic Typology and Linearization. Tech. Rep. **58**(3), 639-645 (1982). https://doi.org/10.2307/413851
15. Laughlin, S.B., de Ruyter van Steveninck, R.R., Anderson, J.C.: The metabolic cost of neural information. Nat. Neurosci. **1**(1), 36–41 (1998). https://doi.org/10.1038/236, http://www.nature.com/articles/nn0598_36
16. Letzkus, J., Wolff, S., Lüthi, A.: Disinhibition, a circuit mechanism for associative learning and memory. Neuron **88**(2), 264–276 (2015). https://doi.org/10.1016/j.neuron.2015.09.024, https://www.sciencedirect.com/science/article/pii/S0896627315008132
17. Maurits, L.: Representation, information theory and basic word order. Tech. Rep. (2011)
18. Maurits, L., Griffiths, T.L.: Tracing the roots of syntax with Bayesian phylogenetics. In: Proceedings of the National Academy of Sciences of the United States of America, vol 111, no. 37, pp. 13576–13581 (2014). https://doi.org/10.1073/pnas.1319042111. https://pubmed.ncbi.nlm.nih.gov/25192934/
19. Maurits, L., Perfors, A., Navarro, D.: Why are some word orders more common than others? A Uniform Information Density account. Tech. rep., vol. 23, pp. 1585–1593(2010)
20. Papadimitriou, C.H., Vempala, S.S., Mitropolsky, D., Collins, M., Maass, W.: Brain computation by assemblies of neurons. In: Proceedings of the National Academy of Sciences **117**(25), 14464–14472 (2020). https://doi.org/10.1073/pnas.2001893117, https://www.pnas.org/content/117/25/14464

21. Papadimitriou, C.H., Yannakakis, M.: The traveling salesman problem with distances one and two. Math. Oper. Res. **18**(1), 1–11 (1993). http://www.jstor.org/stable/3690150
22. Tomlin, R.S.: Basic word order. Functional principles. No. 1, Cambridge University Press (CUP) (1988). https://doi.org/10.1017/s0022226700011646. https://www.cambridge.org/core/journals/journal-of-linguistics/article/abs/russell-s-tomlin-basic-word-order-functional-principles-london-croom-helm-1986-pp-308/7542AFB4A8B28D651F6E109B810F4C04